An acoustic-phonetic account of phonotactic perceptual assimilation

Eleanor Chodroff, Anthony Arnette, Samhita Ilango, and Colin Wilson
Department of Cognitive Science, Johns Hopkins University

167th Meeting of the Acoustical Society of America

Background

Previous research has identified a coronal-to-dorsal 'perceptual assimilation' in which English listeners identify Modern Hebrew word-initial /tl/ and /dl/ as beginning with /k/ and /g/, respectively (Hallé & Best 2007). Reported findings indicate that /tl/ is misperceived more often than /dl/—a surprising asymmetry on phonological grounds—and acoustic-phonetic factors that modulate misperception rates *across stimulus types and tokens* have not been identified.

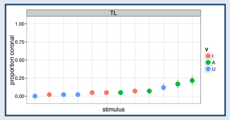
Cross-language perceptual assimilation in general can be attributed to both phonological constraints and processes (e.g., Berent et al. 1997) and acoustic-phonetic (auditory) similarity to native categories (e.g., Escudero et al. 2012).

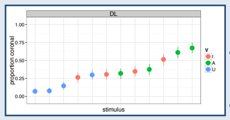
- Phonotactic constraints of English support perceptual repair of/tl/ and /dl/, but do not account for different rates of perceptual assimilation across types and tokens.
- Can acoustic-phonetic properties known to be relevant to place perception account for detailed patterns of perceptual assimilation?

Objectives

- Replicate Hallé & Best study with original stimuli and novel stimuli from a different MH talker
- Determine the role of acoustic-phonetics in coronal-to-dorsal perceptual assimilation

Identification Results: Stimulus Set 1





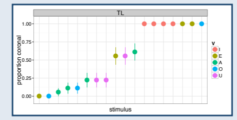
pre-/l/		Response					
		T	D	K	G	P	B
Consonant	T	**32**	4	**434**	19	8	1
	D	1	**165**	2	**309**	1	15
	K	1	0	245	0	1	0
	G	1	2	0	226	1	0

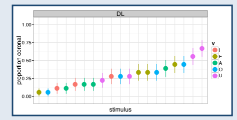
Overall accuracy
pre-/l/ : 45.5%
pre-/ʁ/ : 97.7%

Analyzed with mixed-effects logistic regression, predicting **place perception accuracy** of the pre-/l/ subset.

- Coronal-to-dorsal perceptual assimilation
  (place = -4.45, $p < .001$)
- Replication of /dl/>/tl/ asymmetry
  (place × voice = -0.88, $p < .05$)

Identification Results: Stimulus Set 2





pre-/l/		Response					
		T	D	K	G	P	B
Consonant	T	**173**	1	**138**	1	10	1
	D	8	**89**	11	**174**	14	46
	K	3	0	356	0	1	0
	G	1	1	11	282	2	9

Overall accuracy
pre-/l/ : 67.6%
pre-/ʁ/ : 84.5%*

*primarily voicing errors

Analyzed with mixed-effects logistic regression, predicting **place perception accuracy** of pre-/l/ and pre-/ʁ/ contexts.

- Coronal-to-dorsal perceptual assimilation
  (place = -1.86, $p < .001$; place × C2 = -1.87, $p < .001$)
- More accurate perception of **voiceless** stops
  (voice = 0.91, $p < .01$; place × voice = 0.01, $p = .96$)

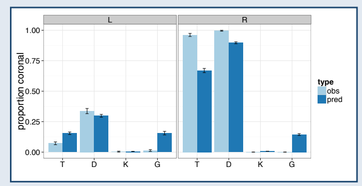
Acoustic Analysis

- Measures to predict rate of coronal response

Measure	Method
• Mean frequency • Spectral s.d. • Skew • Kurtosis	Spectral moments (Forrest et al. 1988) from a smoothed spectrum (Hanson & Stevens 2003; Flemming 2007) of initial ~10ms burst onset
Burst duration	Burst onset to onset of voicing (periodicity)
Relative amplitude	Max amp of initial 10ms of burst – max amp vowel
F3–F2	Measured at approximant onset

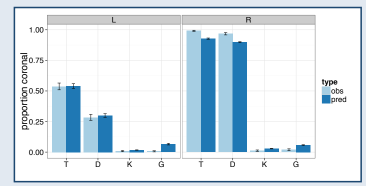
Acoustic Model of Coronal Perception

Stimulus Set 1



Significant factors: mean frequency = 2.39, s.d. = -1.22, skew = -0.86, kurtosis = 1.79, vcl/vot = -1.63, vcd/vot = -2.22, F3–F2 = 1.10 ($ps < .001$)
$r = 0.88$ across stimuli

Stimulus Set 2



Significant factors: mean frequency = 2.83, s.d. = -1.15, kurtosis = 0.72, vcl/vot = -1.18, F3–F2 = 0.41 ($ps < .001$)
$r = 0.93$ across stimuli

Perception Methods

	Stimulus Set 1 (Hallé & Best 2007)	Stimulus Set 2
Stimuli	• male native Hebrew talker • [t k d g] × [l ʁ] × [a i u] • 4 recordings per syllable type (96)	• female native Hebrew talker • [t k d g] × [l ʁ] × [a e i o u] in carrier phrase '/tagit __ ʃuv/' ('Say __ again.') • 4 recordings per syllable type (160)
Participants	23 AE speakers from JHU	18 AE speakers from JHU
Task	In each trial, syllable stimulus presented twice auditorily. Participants identified the initial sound of the word in a 6 alternative forced-choice task [P B T D K G].	

Discussion

Perception

- English listeners' perception of MH /tl/-/dl/ clusters is highly talker- and stimulus-dependent
  - Misperception rates vary from 0% to 100% for /tl/ stimuli
  - /tl/ vs. /dl/ asymmetry is not consistent across talkers
- Main sources of variation in perception, including voiceless vs. voiced asymmetries, are acoustic-phonetic properties

Acoustic Model

- Acoustic-phonetic factors known to cue stop consonant place of articulation can model the rate of coronal identification
- Directions of acoustic effects are expected from the general theory of place perception, for example:
  - Higher mean burst frequency → more coronal responses
    (Set 1: /tl/ 3196 Hz, /dl/ 2713 Hz; Set 2: /tl/ 5133 Hz, /dl/ 4037 Hz)
  - More 'peaked' spectral distrib. → more coronal responses
  - Longer burst duration → fewer coronal responses

Future Directions

- Can perceptual models trained on English stop acoustics predict cross-language perception patterns?
- How much do phonotactic constraints contribute to the perception of illegal consonant clusters?
- Talker adaptation and learning effects? (but no evidence so far for adaptation over the course of an experiment)